Quaderni di Statistica e Matematica
applicata alle Scienze Economico-Sociali
VOLUME IV DICEMBRE 1981

SUI MAGGIORANTI DELLA DISTRIBUZIONE ASINTOTICA DI UN INDICE DI CONCENTRAZIONE FRA LE CLASSI DI GINI*

Michele ZENGA - Antonio BRUNAZZO**

0. Sommario

Nelle indagini socio-economiche, spesso si hanno problemi che riguardano la difforme distribuzione di una quantità trasferibile fra le k classi in cui può essere divisa una popolazione. Una misura della concentrazione fra le classi è data dal noto indice di concentrazione di Gini:

$$\Delta = \frac{D_r}{2\bar{X}}$$

in cui D_r indica la differenza media assoluta con ripetizione fra le medie delle classi e $\bar{X}$ indica, come al solito, la media artimetica generale della quantità trasferibile. Con semplici passaggi è facile esprimere l'indice di concentrazione fra le classi anche nel seguente modo:

$$\Delta = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}|p_i q_j - p_j q_i|$$

in cui:

i) p_i indica la frequenza relativa delle unità appartenenti alla classe $i^{\underline{ma}}$;

ii) q_i indica la quota del carattere trasferibile spettante alla classe $i^{\underline{ma}}$.

Nell'ipotesi in cui sono noti i valori di q_i della popolazione e si disponga di un campione bernoulliano che permetta di stimare le frequenze relative p_i

---

* Lavoro presentato, in forma ristretta, alla XXXI riunione scientifica della Società Italiana di Statistica, Torino 5-7 aprile 1982. I capitoli 0,1,2,3 sono dovuti a M. Zenga mentre i capitoli 4,5,6 e 7 sono dovuti a A. Brunazzo.

** Facoltà di Economia e Commercio, Libera Università degli Studi di Trento

della popolazione attraverso le frequenze $\hat{p}_i$ del campione, è possibile misurare la concentrazione campionaria fra le classi con la statistica:

$$G = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} | \hat{p}_i q_j - \hat{p}_j q_i |$$

La distribuzione asintotica di G è stata ricavata da Colombi [1] nel caso $\Delta = 0$ e $k = 3$ e da Brunazzo [2] nel caso $\Delta = 0$, $k = 4$ e $q_1 = q_2 = q_3 = q_4 = \frac{1}{4}$.

In questo lavoro, sempre per l'ipotesi $\Delta = 0$, si ricavano con due diverse procedure, valori maggioranti dei quantili di G. La prima procedura utilizza la classica disuguaglianza di Bonferroni, l'altra utilizza la distribuzione della v.c. Chi-quadrato centrale.

Le due procedure, in qualche modo, si compensano nel senso che, per valori piccoli di k, i maggioranti che si ricavano con la disuguaglianza di Bonferroni sono minori degli altri ed all'aumentare di k le posizioni si invertono.

Il grado di approssimazione fra i maggioranti qui proposti ed i valori esatti dei quantili trovati da Colombi e da Brunazzo per il caso $k = 3$ e $k = 4$ è del tutto soddisfacente e giustifica l'impiego dei maggioranti in problemi di inferenza statistica.

1. Introduzione e simbologia

Si indichi con X l'intensità di un carattere trasferibile di cui si vuole misurare la concentrazione fra le k classi in cui è divisa la popolazione. Sia N_i il numero delle unità appartenenti alla classe $i^{\underline{ma}}$ $(i = 1,2,\ldots, k)$ e sia X_{ij} l'intensità della generica $j^{\underline{ma}}$ unità della $i^{\underline{ma}}$ classe $(j = 1,2,\ldots, N_i)$.

E' possibile, così calcolare le seguenti grandezze statistiche:

i) $$p_i = \frac{N_i}{\sum_{i=1}^{k} N_i}$$
$$= \frac{N_i}{N} \,,$$ frequenza relativa delle unità appartenenti alla classe $i^{\underline{ma}}$.

ii) $\bar{X}_i = \dfrac{\sum_{j=1}^{N_i} X_{ij}}{N_i}$, media aritmetica del carattere nella classe $i^{\underline{ma}}$;

iii) $\bar{X} = \dfrac{\sum_{i=1}^{k} N_i \bar{X}_i}{\sum_{i=1}^{k} N_i}$, media artimetica generale;

iV) $Q_i = N_i \bar{X}_i$, quantità del carattere appartenente alla classe $i^{\underline{ma}}$;

V) $q_i = \dfrac{Q_i}{\sum_{i=1}^{k} Q_i}$

$= \dfrac{N_i \cdot \bar{X}_i}{N \bar{X}}$ quota del carattere appartenente alla classe $i^{\underline{ma}}$.

Per misurare la concentrazione fra le classi si può usare l'indice di concentrazione di Gini:

$$\Delta = \frac{D_r}{2\bar{X}}$$

in cui D_r indica la differenza media assoluta con ripetizione fra le medie $\bar{X}_i$ delle classi. La concentrazione fra le classi, si può, esplicitando l'espressione di D_r, riscrivere come segue:

$$\Delta = \frac{1}{2\bar{X}} \sum_{i=1}^{k} \sum_{j=1}^{k} \left| \bar{X}_j - \bar{X}_i \right| = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \left| \frac{\bar{X}_j}{\bar{X}} - \frac{\bar{X}_i}{\bar{X}} \right| p_i p_j \ .$$

Tenuto presente che $\dfrac{\bar{X}_i}{\bar{X}} = \dfrac{q_i}{p_i}$, l'espressione precedente diventa:

$$\Delta = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \left| p_i q_j - p_j q_i \right| \ . \qquad (1.1)$$

Nell'ipotesi di equiripartizione si ha:

$$\Delta = 0 \leftrightarrow p_i = q_i \quad (i = 1,2,\ldots, k).$$

Si estragga, con riposizione, dalla popolazione un campione casuale di ampiezza n e si indichi n_i il numero delle unità estratte appartenenti alla classe $i^{\underline{ma}}$ (ovviamente: $n_i > 0; i = 1,2,\ldots, k; \sum n_i = n$). Le frequenze $(n_1, n_2, \ldots n_k)$ costituiscono la determinazione campionaria di una distribuzione multinomiale $(n; p_1, p_2, \ldots, p_k)$. Nell'ipotesi in cui siano noti i valori q_i della popolazione, dai dati campionari è possibile calcolare l'indice di Gini che assumerà la forma:

$$G = \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \left| \frac{n_i}{n} q_j - \frac{n_j}{n} q_i \right| \tag{1.2}$$

Nei paragrafi che seguono si indagherà sulla distribuzione campionaria della (1.2) nell'ipotesi che $\Delta = 0$.

2. Impiego della disuguaglianza di Bonferroni, caso k = 3

In questo paragrafo verrá esaminato, in dettaglio, il procedimento che ha permesso di ricavare i maggioranti con l'uso della disuguaglianza di Bonferroni. Si è preferito esaminare a parte il caso k = 3 in quanto la trattazione per un k generico risulta abbastanza complessa, come si noterà nel paragrafo 4 e seguenti. Con semplici passaggi è possibile descrivere la (1.2) come segue:

$$G = \frac{1}{n} \left\{ q_1 q_2 \left| \frac{n_1}{q_1} - \frac{n_2}{q_2} \right| + q_1 q_3 \left| \frac{n_1}{q_1} - \frac{n_3}{q_3} \right| + q_2 q_3 \left| \frac{n_2}{q_2} - \frac{n_2}{q_3} \right| \right\}. \tag{2.1}$$

Si considerino ora le seguenti sei regioni R_i, $(i = 1,2,\ldots, 3!)$ dello spazio R^2:

$$R_1 = \left\{ (n_1, n_2, n_3) : \frac{n_3}{q_3} \leqslant \frac{n_2}{q_2} \leqslant \frac{n_1}{q_1} \right\}$$

$$R_2 = \left\{ (n_1, n_2, n_3) : \frac{n_1}{q_1} \leqslant \frac{n_2}{q_2} \leqslant \frac{n_3}{q_3} \right\}$$

$$R_3 = \left\{ (n_1, n_2, n_3) : \frac{n_3}{q_3} \leqslant \frac{n_1}{q_1} \leqslant \frac{n_2}{q_2} \right\}$$

$$R_4 = \left\{ (n_1, n_2, n_3) : \frac{n_2}{q_2} \leqslant \frac{n_1}{q_1} \leqslant \frac{n_3}{q_3} \right\}$$

$$R_5 = \left\{(n_1, n_2, n_3) : \frac{n_2}{q_2} \leqslant \frac{n_3}{q_3} \leqslant \frac{n_1}{q_1}\right\}$$

$$R_6 = \left\{(n_1, n_2, n_3) : \frac{n_1}{q_1} \leqslant \frac{n_3}{q_3} \leqslant \frac{n_2}{q_2}\right\}$$

Si indichi con G_i l'espressione con cui si può scrivere la (2.1), senza ricorrere ai valori assoluti, nell'ipotesi $(n_1, n_2, n_3) \in R_i$.

In particolare si ha:

$$G_1 = \frac{1}{n}\left\{q_1 q_2 \left(\frac{n_1}{q_1} - \frac{n_2}{q_2}\right) + q_i q_3 \left(\frac{n_1}{q_1} - \frac{n_3}{q_3}\right) + q_2 q_3 \left(\frac{n_2}{q_2} - \frac{n_3}{q_3}\right)\right\};$$

$$G_2 = - G_1 ;$$

$$G_3 = \frac{1}{n}\left\{- q_1 q_2 \left(\frac{n_1}{q_2} - \frac{n_2}{q_2}\right) + q_1 q_3 \left(\frac{n_1}{q_1} - \frac{n_3}{q_3}\right) + q_2 q_3 \left(\frac{n_2}{q_2} - \frac{n_3}{q_3}\right)\right\};$$

$$G_4 = - G_3 ;$$

$$G_5 = \frac{1}{n}\left\{q_1 q_2 \left(\frac{n_1}{q_1} - \frac{n_2}{q_2}\right) + q_1 q_3 \left(\frac{n_1}{q_1} - \frac{n_3}{q_3}\right) - q_2 q_3 \left(\frac{n_2}{q_2} - \frac{n_3}{q_3}\right)\right\};$$

$$G_6 = - G_5 .$$

Riordinando i termini delle espressioni precedenti e ricordando che $n_3 = n - n_1 - n_2$, si ha ancora:

$$G_1 = \frac{1}{n}\left\{n_1 (1 + q_2) + n_2 (1 - q_1) - n (1 - q_3)\right\} ; \qquad (2.2)$$

$$G_3 = \frac{1}{n}\left\{n_1 (1 - q_2) + n_2 (1 + q_1) - n (1 - q_3)\right\} ; \qquad (2.3)$$

$$G_5 = \frac{1}{n}\left\{n_1 (1 - q_2) - n_2 (1 - q_1) - n (q_1 - q_2)\right\}. \qquad (2.4)$$

Ovviamente è possibile considerare la G_i anche fuori dalla regione R_i, però il suo valore, in questo caso, non coincide con quello di G.
Si consideri ora il luogo dei punti dello spazio campionario R^2 in cui $G_1 = g$, $(g > 0)$: trattasi dei punti caratterizzati dalla retta di equazione:

$$n_1 (1 + q_2) + n_2 (1 - q_1) - n (1 - q_3) = ng.$$

Inoltre il luogo dei punti in cui $G_1 = -g$ è la retta, parallela alla precedente, di equazione:

$$n_1 (1 + q_2) + n_2 (1 - q_1) - n (1 - q_3) = - ng.$$

Indichiamo con A_1 l'evento: $-g \leq G_1 \leq g$. In maniera analoga è possibile caratterizzare l'evento $A_2 : -g \leq G_3 \leq g$ e l'evento $A_3 : -g \leq G_5 \leq g$. Risulta altresì evidente che l'evento $G \leq g$ è equivalente all'evento $\{A_1 \cap A_2 \cap A_3\}$ e pertanto $G > g$ è equivalente all'evento $\{\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3\}$.
La disuguaglianza di Bonferroni afferma che:

$$P\{\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3\} \leq P[\bar{A}_1] + P[\bar{A}_2] + P[\bar{A}_3]. \tag{2.5}$$

L'utilizzazione di questa disuguaglianza ci permette di trovare, nel caso asintotico, un valore maggiorante di g_α, essendo

$$P[G > g_\alpha \mid \Delta = 0] = \alpha . \tag{2.6}$$

Indichiamo con ${}^+g_\alpha$ il maggiorante di g_α, che verrà ricavato in questo paragrafo (ovviamente ${}^+g_\alpha > g_\alpha$ e quindi $P[G > {}^+g_\alpha \mid \Delta = 0] \leq \alpha$).
Sotto l'ipotesi $\Delta = 0$ le frequenze campionarie (n_1, n_2, n_3) sono distribuite secondo una trinomiale con parametri $(n; p_1 = q_1, p_2 = q_2, p_3 = q_3)$ e pertanto:

$$E(n_i \mid \Delta = 0) = nq_i \qquad (i = 1,2,3) \tag{2.7}$$

$$\mathrm{var}(n_i \mid \Delta = 0) = nq_i (1-q_i) \qquad (i = 1,2,3) \tag{2.8}$$

$$\mathrm{cov}(n_i n_j \mid \Delta = 0) = -nq_i q_j \qquad (i, j = 1,2,3; i \neq j) . \tag{2.9}$$

E' facile constatare che:

$$E(G_i \mid \Delta = 0) = 0, \qquad (i = 1,2,3) \tag{2.10}$$

$$\mathrm{var}(G_i \mid \Delta = 0) = \frac{1}{n}(1-q_1)(1-q_2)(1-q_3), \; (i = 1,2,3) . \tag{2.11}$$

Al divergente di n, e sotto l'ipotesi $\Delta = 0$, è possibile approssimare la v.c.

$$Y_i = \frac{\sqrt{n}\, G_i}{\sqrt{(1-q_1)(1-q_2)(1-q_3)}} \quad , \ (i = 1,2,3) \tag{2.12}$$

con una v.c. normale standardizzata. Si indichi ora, come di consueto, con $\Phi(z)$ la funzione cumulata di una distribuzione normale standardizzata e si ponga:

$$^{+}y_{\alpha} = \Phi^{-1}\left(1 - \frac{\alpha}{3!}\right) \ , \tag{2.13}$$

ne segue che:

$$P\left[\, |Y_i| > {}^{+}y_{\alpha} \,\right] = \frac{\alpha}{3} \tag{2.14}$$

e quindi:

$$P\left[\, (|Y_1| \geqslant {}^{+}y_{\alpha}) \cup (|Y_2| \geqslant {}^{+}y_{\alpha}) \cup (|Y_3| \geqslant {}^{+}y_{\alpha}) \leqslant \alpha \right. \ . \tag{2.15}$$

Dalla (2.12) si ricava:

$$G_i = \frac{Y_i}{\sqrt{n}} \sqrt{(1-q_1)(1-q_2)(1-q_3)}$$

e quindi, dalla (2.13),

$$P\left[\bigcup_{i=1}^{3} |G_i| \geqslant {}^{+}y_{\alpha}\, \frac{\sqrt{(1-q_1)(1-q_2)(1-q_3)}}{\sqrt{n}}\right] \leqslant \alpha \ . \tag{2.16}$$

Conseguentemente il maggiorante che si ricava con questo procedimento è:

$$^{+}g_{\alpha} = \frac{\sqrt{(1-q_1)(1-q_2)(1-q_3)}}{\sqrt{n}}\, \Phi^{-1}\left(1 - \frac{\alpha}{3!}\right) \tag{2.17}$$

3. Impiego della distribuzione chi-quadrato, caso k = 3

Anche per questo procedimento viene esaminato a parte il caso $k = 3$.
Sotto l'ipotesi $\Delta = 0$, la v.c.

$$X^2 = \frac{(n_1 - nq_1)^2}{nq_1} + \frac{(n_2 - nq_2)^2}{nq_2} + \frac{(n - n_1 - n_2 - nq_3)^2}{nq_3} \quad (3.1)$$

si distribuisce, al divergere di n, secondo una v.c. chi-quadrato centrale con 2 gradi di libertà.
L'evento A_1 definito nel secondo paragrafo è rappresentato, in R^2, dalla parte di piano compresa fra le due rette parallele:

$$ng = n_1 (1 + q_2) + n_2 (1 - q_1) - n (1 - q_3) \quad (3.2)$$

$$-ng = n_1 (1 + q_2) + n_2 (1 - q_1) - n (1 - q_3). \quad (3.3)$$

In maniera analoga saranno rappresentati in R^2 gli eventi A_2 e A_3.
L'intersezione delle parti di piano individuate dagli eventi A_1, A_2 e A_3 dà luogo ad un esagono $E_g = \bigcap_{i=1}^{3} A_i$, con i lati giacenti su equazioni del tipo della (3.2) e della (3.3). L'esagono E_g è caratterizzato dal fatto che la v.c. G assume in esso valori al più uguali a g. Dato che per ogni assegnato valore positivo assunto da X^2 la (3.1) rappresenta, in R^2, una ellisse, ci proponiamo di determinare, fra tutte queste ellissi, la più grande contenuta in E_g. A tal fine basterà minimizzare la (3.1) sui lati dell'esagono E_g. Consideriamo, ad esempio, il lato di E_g che si trova sulla retta (3.2). Come di consueto usiamo il procedimento dei moltiplicatori di Lagrange:

$$F(n_1, n_2, \lambda) = X^2 + \lambda [n_1 (1 + q_2) + n_2 (1 - q_1) - n (1 - q_3) - ng] \quad (3.4)$$

$$\frac{\partial F}{\partial n_1} = \frac{2(n_1 - nq_1)}{nq_1} - \frac{2(n - n_1 - n_2 - nq_3)}{nq_3} + \lambda (1 + q_2)$$

$$\frac{\partial F}{\partial n_2} = \frac{2(n_2 - nq_2)}{nq_2} - \frac{2(n - n_1 - n_2 - nq_3)}{nq_3} + \lambda(1 - q_1)$$

$$\frac{\partial F}{\partial \lambda} = n_1 (1 + q_2) + n_2 (1 - q_1) - n(1 - q_3) - ng.$$

Uguagliate a zero le derivate parziali, riordinati i termini e tenuto presente che $n_3 = n - n_1 - n_2$, si ottiene, dopo alcuni passaggi, la soluzione:

$$\hat{n}_1 = nq_1 + \frac{g}{d} nq_1 (1 - q_1) \tag{3.5}$$

$$\hat{n}_2 = nq_2 + \frac{g}{d} nq_2 (q_3 - q_1) \tag{3.6}$$

$$\hat{n}_3 = nq_3 - \frac{g}{d} nq_3 (1 - q_3) \tag{3.7}$$

dove: $$d = (1 - q_1)(1 - q_2)(1 - q_3). \tag{3.8}$$

Calcolando il valore di X^2 in corrispondenza del punto $(\hat{n}_1, \hat{n}_2, \hat{n}_3)$ si ha:

$$X^2 = \frac{n}{d} g^2 . \tag{3.9}$$

Siccome il risultato trovato non varia sostituendo, nel procedimento, il lato che si trova su (3.2) con un qualsiasi altro lato di E_g possiamo concludere che esiste un'ellisse tangente ai lati dell'esagono E_g, caratterizzata dalla (3.1) in cui X^2 é sostituito con $\frac{n}{d} g^2$.

Indicando con $\chi^2_{(1-\alpha),2}$ il $(1-\alpha)$quantile della v.c. chi-quadrato con 2 gradi di libertà è immediato notare che:

$$g^+_\alpha = \sqrt{\frac{d}{n}} \sqrt{\chi^2_{(1-\alpha),2}} \tag{3.10}$$

è il cercato maggiorante di g_α; infatti g^+_α è tale che:

$$P[G \geqslant g^+_\alpha \mid \Delta = 0] \leqslant \alpha . \tag{3.11}$$

Si nota che il maggiorante (3.10) differisce da quello dato dalla (2.17) perchè, in quest'ultima, al posto di $\Phi^{-1}\left(1 - \frac{\alpha}{3!}\right)$ appare $\sqrt{\chi^2_{(1-\alpha),2}}$.

4. Impiego della disuguaglianza di Bonferroni. Caso generale.

In questo paragrafo, per l'espressione (1.2), useremo la seguente simbologia:

$$G = \sum_{\substack{i,j=1 \\ i<j}}^{k} \left| \frac{n_i}{n} q_j - \frac{n_j}{n} q_i \right| \tag{4.1}$$

dove occorre ricordare che le quantità q_i $(i = 1,2,\ldots,k)$ sono note. Ciò premesso, si considerino, nello spazio R^{k-1}, le k! regioni così definite:

$$R_{i_1 i_2 \ldots i_k} = \left\{ (n_1, n_2, \ldots, n_k) : \frac{n_{i_1}}{q_{i_1}} \leqslant \frac{n_{i_2}}{q_{i_2}} \leqslant \ldots \leqslant \frac{n_{i_k}}{q_{i_k}} \right\}, \tag{4.2}$$

essendo la k-pla $(i_1 i_2 \ldots i_k)$ una qualsiasi permutazione dei k numeri $(1,2,\ldots k)$. Nell'ipotesi che $(n_1, n_2, \ldots, n_k) \in R_{i_1 i_2 \ldots i_k}$, è possibile ricavare una funzione $G_{i_1 i_2 \ldots i_k}$ tale che assume gli stessi valori di G (se: $(n_1, n_2, \ldots n_k) \in R_{i_1 i_2 \ldots i_k}$) e che non fa uso dei moduli, precisamente sia:

$$G_{i_1 i_2 \ldots i_k} = \sum_{\substack{i,j=1 \\ i<j}}^{k} \delta_{ij} \left(\frac{n_i}{n} q_j - \frac{n_j}{n} q_i \right), \tag{4.3}$$

dove:

$$\delta_{ij} = \begin{cases} -1, \text{ se i precede j nella k-pla } (i_1 i_2 \ldots i_k) \\ +1, \text{ se i segue j nella k-pla } (i_1 i_2 \ldots i_k); \end{cases} \tag{4.4}$$

sarà: $\delta_{ij} + \delta_{ji} = 0$ e la (4.3) si può scrivere:

$$n\, G_{i_1 i_2 \ldots i_k} = \sum_{\substack{i,j=1 \\ i<j}}^{k} \delta_{ij}\, n_i q_i + \sum_{\substack{i,j=1 \\ i<j}}^{k} \delta_{ji}\, n_j q_i \quad . \tag{4.5}$$

Si indichi inoltre con K la classe di tutte le k! permutazioni dei k numeri $(1,2,\ldots, k)$. Dato che in K ad ogni permutazione possiamo far corrispondere quella che ha gli elementi disposti in ordine inverso, cioè:

$$(i_1, i_2, \ldots, i_k) \rightleftarrows (i_k, i_{k-1}, \ldots, i_1),$$

consideriamo una partizione di K in due classi K_1, K_2 in modo che tutte le permutazioni che sono in K_1 (K_2) abbiano la corrispondente permutazione in K_2 (K_1). Ovviamente ognuna delle due classi possiede $\frac{1}{2}k!$ permutazioni e ne segue che:

$$G_{i_1 i_2 \ldots i_k} = - G_{i_k i_{k-1} \ldots i_1}, \ \forall (i_1 i_2 \ldots i_k) \varepsilon K. \tag{4.6}$$

Lo sviluppo poi di (4.5) è perciò di (4.3) si può evidenziare considerando la matrice quadrata di ordine k:

$$M_{i_1 i_2 \ldots i_k} = \begin{bmatrix} 0 & \delta_{12} n_1 q_2 & \delta_{13} n_1 q_3 \ldots\ldots\ldots & \delta_{1k} n_1 q_k \\ \delta_{21} n_2 q_1 & 0 & \delta_{23} n_2 q_3 \ldots\ldots\ldots & \delta_{2k} n_2 q_k \\ & & & \\ ------ & ------ & ------ & ------ \\ \delta_{k1} n_k q_1 & \delta_{k2} n_k q_2 & \delta_{k3} n_k q_3 \ldots\ldots\ldots\ldots & 0 \end{bmatrix} \tag{4.7}$$

ove si può notare che, in essa, tutti gli elementi (esclusi quelli della diagonale principale) rappresentano i termini dello sviluppo (4.5); infatti per $i < j$:

$\delta_{ij} n_i q_j$ rappresentano gli elementi di M che si trovano a destra della diagonale principale

$\delta_{ji} n_j q_i$ rappresentano gli elementi di M che si trovano a sinistra della diagonale principale.

Osservando la (4.7) si deduce che la (4.5) si può scrivere:

$$\begin{aligned} n G_{i_1 i_2 \ldots i_k} = \ & n_1 (\delta_{12} q_2 + \delta_{13} q_3 + \ldots + \delta_{1k} q_k) + \\ & n_2 (\delta_{21} q_1 + \delta_{23} q_3 + \ldots + \delta_{2k} q_k) + \\ & ------------------ \\ & n_k (\delta_{k1} q_1 + \delta_{k2} q_2 + \ldots + \delta_{kk-1} q_{k-1}), \end{aligned} \tag{4.8}$$

e detti:

$$\alpha_1 = \delta_{12} q_2 + \delta_{13} q_3 + \ldots\ldots + \delta_{1k} q_k ,$$

$$\alpha_2 = \delta_{21} q_1 + \delta_{23} q_3 + \ldots\ldots + \delta_{2k} q_k ,$$

$$- -$$

$$\alpha_k = \delta_{k1} q_1 + \delta_{k2} q_2 + \ldots\ldots + \delta_{k\,k-1} q_{k-1} ,$$

avremo,

$$n\, G_{i_1 i_2 \ldots i_k} = \sum_{i=1}^{k} \alpha_i n_i, \quad \forall\, (i_1 i_2 \ldots i_k) \in K. \tag{4.9}$$

Sostituendo n_k con $n - \sum_{i=1}^{k-1} n_i$ nella (4.9) otteniamo:

$$n\, G_{i_1 i_2 \ldots i_k} = \sum_{i=1}^{k-1} (\alpha_i - \alpha_k)\, n_i + n\, \alpha_k, \ \forall\, (i_1 i_2 \ldots i_k) \in K \,, \tag{4.10}$$

espressione lineare nelle variabili $n_1 , n_2 , \ldots n_{k-1}$.
Ovviamente è possibile considerare la (4.9) e quindi la (4.10) fuori dalla corrispondente $R_{i_1 i_2 \ldots i_k}$ in cui è stata definita, ma in tal caso il valore $G_{i_1 i_2 \ldots i_k}$ non coincide con G.

Si consideri ora, in R^{k-1}, il luogo dei punti per cui:

$$G_{i_1 i_2 \ldots i_k} = g, \ (g > 0) \,, \quad \forall\, (i_1 i_2 \ldots i_k) \in K_1 ,$$

trattasi dei punti caratterizzati dalla $\frac{1}{2}$ k! equazioni:

$$ng = \sum_{i=1}^{k-1} (\alpha_i - \alpha_k)\, n_i + n\, \alpha_k , \tag{4.11}$$

rappresentanti ciascuna un iperpiano. Inoltre il luogo dei punti per cui $G_{i_1 i_2 \ldots i_k} = -g, \ \forall\, (i_1 i_2 \ldots i_k) \in K_1$, sono iperpiani paralleli ai precedenti

di equazione:

$$-ng = \sum_{i=1}^{k-1} (\alpha_i - \alpha_k)\, n_i + n\, \alpha_k \,, \ \forall\, (i_1 i_2 \ldots i_k) \in K_1 \tag{4.12}$$

Si indichi con $A_{i_1 i_2 \dots i_k}$ l'evento:

$$-g \leqslant G_{i_1 i_2 \dots i_k} \leqslant g, \forall (i_1 i_2 \dots i_k) \varepsilon K_1 ; \tag{4.13}$$

la loro intersezione:

$$\{\bigcap A_{i_1 i_2 \dots i_k}\} \tag{4.14}$$

è equivalente all'evento:

$$G \leqslant g , \tag{4.15}$$

mentre l'evento $G > g$ è equivalente all'evento:

$$\{\bigcup \bar{A}_{i_1 i_2 \dots i_k}\}. \tag{4.16}$$

Sapendo che, per la disuguaglianza di Bonferroni,

$$P\{\bigcup \bar{A}_{i_1 i_2 \dots i_k}\} \leqslant \sum P(\bar{A}_{i_1 i_2 \dots i_k}), \tag{4.17}$$

l'utilizzazione di (4.17) ci permette di trovare, nel caso asintotico, un valore maggiorante di g_α, essendo quest'ultimo così definito:

$$P[G > g_\alpha | \Delta = 0] = \alpha . \tag{4.18}$$

Indicheremo con ${}^{+}g_{\alpha,k}$ il maggiorante di g_α, cioè

$${}^{+}g_{\alpha,k} \geqslant g_\alpha , \tag{4.19}$$

da cui:

$$P[G \geqslant {}^{+}g_{\alpha,k} | \Delta = 0] \leqslant \alpha . \tag{4.20}$$

Sotto l'ipotesi di $\Delta = 0$, le frequenze campionarie $(n_1, n_2, \dots n_k)$ sono distribuite secondo una multinomiale di parametri: $(n; p_1 = q_1, p_2 = q_2, \dots, p_k = q_k)$ per cui:

$$\begin{array}{ll} E(n_i \mid \Delta = 0) = nq_i & (i = 1,2,3,\dots, k), \\ \mathrm{var}(n_i \mid \Delta = 0) = nq_i(1-q_i) & (i = 1,2,3,\dots, k), \\ \mathrm{cov}(n_i n_j | \Delta = 0) = -nq_i q_j & (i, j = 1,2,\dots, k; i \neq j), \end{array}$$

e inoltre, come sarà dimostrato in seguito, $\forall\ (i_1 i_2 \dots i_k) \varepsilon K$:

$$E\,(G_{i_1 i_2 \dots i_k} \mid \Delta = 0) = 0,$$

$$\operatorname{var}\,(G_{i_1 i_2 \dots i_k} \mid \Delta = 0) = \frac{1}{n}\left[2\sum_{\substack{r_1, r_2, r_3 = 1 \\ r_1 < r_2 < r_3}}^{k} q_{r_1} q_{r_2} q_{r_3} + \sum_{\substack{i, j = 1 \\ i<j}}^{k} q_i q_j (q_i + q_j)\right]. \tag{4.21}$$

Ne segue che, al divergere di n e sotto l'ipotesi $\Delta = 0$, è possibile approssimare la v.c.

$$Y_{i_1 i_2 \dots i_k} = \frac{\sqrt{n}}{\sqrt{d}}\, G_{i_1 i_2 \dots i_k} \tag{4.22}$$

con una v.c. normale standardizzata, essendo:

$$d = 2\sum_{\substack{r_1, r_2, r_3 = 1 \\ r_1 < r_2 < r_3}}^{k} q_{r_1} q_{r_2} q_{r_3} + \sum_{\substack{i, j = 1 \\ i<j}}^{k} q_i q_j (q_i + q_j)\ . \tag{4.23}$$

Si indichi con $\Phi(z)$ la funzione cumulata di una distribuzione normale standardizzata e si definisca:

$${}^{+}y_{\alpha,k} = \Phi^{-1}\ (1 - \frac{\alpha}{k!}). \tag{4.24}$$

Conseguentemente:

$$P\,[\,|\,Y_{i_1 i_2 \dots i_k}| > {}^{+}y_{\alpha,k}\,] = \frac{2\alpha}{k!} \tag{4.25}$$

e

$$P[\bigcup |\,Y_{i_1 i_2 \dots i_k}| > {}^{+}y_{\alpha,k}] \leqslant \alpha\ , \tag{4.26}$$

l'unione essendo estesa a tutte le $\frac{1}{2}$ k! permutazioni di K_1.
Dall (4.22) si ricava:

$$G_{i_1 i_2 \dots i_k} = \frac{\sqrt{d}}{\sqrt{n}}\, Y_{i_1 i_2 \dots i_k} \tag{4.27}$$

e quindi:

$$^{+}g_{\alpha,k} = \frac{\sqrt{d}}{\sqrt{n}}\, \Phi^{-1}\left(1 - \frac{\alpha}{k!}\right) \tag{4.28}$$

5. Calcolo di: $E(G_{i_1 i_2 \ldots i_k} \mid \Delta = 0)$, $\text{var}(G_{i_1 i_2 \ldots i_k} \mid \Delta = 0)$

Ricordando la (4.3), estesa a tutto R^{k-1}, cioè:

$$n\, G_{i_1 i_2 \ldots i_k} = \sum_{\substack{i,j=1\\ i<j}}^{k} \delta_{ij}\,(n_i q_j - n_j q_i),$$

si ha

$$E(n\, G_{i_1 i_2 \ldots i_k} \mid \Delta = 0) = E\Big(\sum_{\substack{i,j=1\\ i<j}}^{k} \delta_{ij}(n_i q_j - n_j q_i)\Big) = \sum_{\substack{i,j=1\\ i<j}}^{k} \delta_{ij}(q_j E(n_i) - q_i E(n_j)) =$$

$$= n \sum_{\substack{i,j=1\\ i<j}}^{k} \delta_{ij}\,(q_j q_i - q_i q_j),$$

e quindi:

$$E(G_{i_1 i_2 \ldots i_k} \mid \Delta = 0) = 0\ ,\ \forall\, (i_1 i_2 \ldots i_k)\ \varepsilon K\ . \tag{5.1}$$

Per il calcolo della varianza, utilizzando la (4.9) estesa a tutto R^{k-1}, avremo:

$$\text{var}(n\, G_{i_1 i_2 \ldots i_k} \mid \Delta = 0) = \text{var}\Big(\sum_{i=1}^{k} \alpha_i n_i\Big) = \sum_{i=1}^{k} \text{var}(\alpha_i n_i) + 2 \sum_{\substack{i,j=1\\ i<j}}^{k} \alpha_i \alpha_j \text{cov}(n_i n_j) =$$

$$= n \sum_{i=1}^{k} \alpha_i^2 q_i (1 - q_i) - 2n \sum_{\substack{i,j=1\\ i<j}}^{k} \alpha_i \alpha_j q_i q_j = n \Big[\sum_{i=1}^{k} \alpha_i^2 q_i - \Big(\sum_{i=1}^{k} \alpha_i^2 q_i^2 +$$

$$+ 2 \sum_{\substack{i\,j=1\\ i<j}}^{k} \alpha_i \alpha_j q_i q_j\Big)\Big] = n \Big[\sum_{i=1}^{k} \alpha_i^2 q_i - \Big(\sum_{i=1}^{k} \alpha_i q_i\Big)^2\Big].$$

I termini di $\sum_{i=1}^{k} \alpha_i q_i$ sono gli elementi della matrice **M**, nella quale si è sostituito q_i al posto di n_i, $(i = 1,2, \ldots, k)$, conseguentemente:

$$\sum_{i=1}^{k} \alpha_i q_i = \sum_{\substack{i,j=1 \\ i<j}}^{k} q_i q_j (\delta_{ij} + \delta_{ji}) = 0, \quad \text{essendo: } \delta_{ij} + \delta_{ji} = 0 .$$

Avremo quindi:

$$n \operatorname{var} (G_{i_1 i_2 \ldots i_k} \mid \Delta = 0) = \sum_{i=1}^{k} \alpha_i^2 q_i . \tag{5.2}$$

Dato che:

$$\begin{aligned}\sum_{i=1}^{k} \alpha_i^2 q_i = {} & q_1 (\delta_{12} q_2 + \delta_{13} q_3 + \ldots\ldots + \delta_{1k} q_k)^2 + \\ & q_2 (\delta_{21} q_1 + \delta_{23} q_3 + \ldots\ldots + \delta_{2k} q_k)^2 + \\ & - \\ & q_k (\delta_{k1} q_1 + \delta_{k2} q_2 + \ldots\ldots + \delta_{k\,k-1} q_{k-1})^2 ,\end{aligned}$$

si ottiene:

$$\begin{aligned}\sum_{i=1}^{i} \alpha_i^2 q_i = {} & q_1 (q_2^2 + q_3^2 + \ldots\ldots + q_k^2 + 2\, \delta_{12}\, \delta_{13} q_2 q_3 + \ldots\ldots) + \\ & \underline{+}\, q_2 (q_1^2 + q_2^2 + \ldots\ldots + q_k^2 + 2\, \delta_{21}\, \delta_{23} q_1 q_3 + \ldots\ldots) + \\ & - \\ & + q_k (q_1^2 + q_2^2 + \ldots\ldots + q_{k-1}^2 + 2\, \delta_{k1}\, \delta_{k2}\, q_1 q_2 + \ldots\ldots) .\end{aligned} \tag{5.3}$$

Dalla (5.3) si ha:

a) I termini del tipo $q_r^2\ q_s$, $r \neq s$, si possono evidenziare presentandoli come elementi che, sommati a coppie, forniscono:

$$\sum_{\substack{i,j=1 \\ i<j}}^{k} q_i\, q_j\, (q_i + q_j) . \tag{5.4}$$

b) Il termine nel quale compare il fattore $q_{r_1} q_{r_2} q_{r_3}$ $(r_1 \neq r_2 \neq r_3 ; r_1, r_2, r_3 = 1,2, \ldots, k)$ compare tre volte, precisamente:

$$2q_{r_1} (q_{r_2} q_{r_3} \delta_{r_1 r_2} \delta_{r_1 r_3}) \text{ nella riga: } r_1\text{-ma}$$

$$2q_{r_2} (q_{r_1} q_{r_3} \delta_{r_2 r_1} \delta_{r_2 r_3}) \text{ nella riga: } r_2\text{-ma}$$

$$2q_{r_3} (q_{r_1} q_{r_2} \delta_{r_3 r_1} \delta_{r_3 r_2}) \text{ nella riga: } r_3\text{-ma} .$$

Potremo quindi scrivere la somma di tutti questi prodotti come segue:

$$2\sum_{\substack{r_1, r_2, r_3 = 1 \\ r_1 < r_2 < r_3}}^{k} q_{r_1} q_{r_2} q_{r_3} (\delta_{r_1 r_2} \delta_{r_1 r_2} + \delta_{r_2 r_1} \delta_{r_2 r_3} + \delta_{r_3 r_1} \delta_{r_3 r_2}) \quad (5.5)$$

e dato che: $\delta_{r_1 r_2} \delta_{r_1 r_3} + \delta_{r_2 r_1} \delta_{r_2 r_3} + \delta_{r_3 r_1} \delta_{r_3 r_2} = 1$

qualunque sia l'ordine di precedenza di r_1 con r_2 e di r_2 con r_3, confrontando la terna (r_1, r_2, r_3) con l'allineamento $(i_1 i_2 \ldots i_k)$, potremo concludere:

$$n \operatorname{var} (G_{i_1 i_2 \ldots i_k} \mid J = 0) = 2\sum_{\substack{r_1, r_2, r_3 = 1 \\ r_1 < r_2 < r_3}}^{k} q_{r_1} q_{r_2} q_{r_3} + \sum_{\substack{i, j = 1 \\ i < j}}^{k} q_i q_j (q_i + q_j) . \quad (5.6)$$

Facciamo notare che dalla (5.6) discende immediatamente la (2.11) infatti:

$$n \operatorname{var} (G_{i_1 i_2 i_3} \mid J = 0) = 2\, q_1 q_2 q_3 + \sum_{\substack{i, j = 1 \\ i < j}}^{3} q_i q_j (q_i + q_j) =$$

$$= q_1 q_2 q_3 + q_1 q_2 q_3 + q_1 q_2 (q_1 + q_2) + q_1 q_3 (q_1 + q_3) + q_2 q_3 (q_2 + q_3) =$$

$$= q_1 q_2 (q_3 + q_1 + q_2) + q_1 q_3 (q_2 + q_1 + q_3) + q_2 q_3 (q_2 + q_3) = q_1 q_2 + q_1 q_3 + q_2 q_3 (q_2 + q_3) =$$

$$= q_1 (q_2 + q_3) + q_2 q_3 (q_2 + q_3) = (q_2 + q_3) (q_1 + q_2 q_3) = (q_2 + q_3) [q_1 + q_2 (1 - q_1 - q_2)] =$$

$$= (q_2 + q_3) [q_1 + q_2 - q_1 q_2 - q_2^2] = (q_2 + q_3) [q_1 (1 - q_2) + q_2 (1 - q_2)] =$$

$$= (q_2 + q_3) (1 - q_2) (q_1 + q_2) = (1 - q_1) (1 - q_2) (1 - q_3),$$ che è appunto la (2.11).

In particolare se: $q_1 = q_2 = \ldots = q_k = \frac{1}{k}$, sarà:

$$n \operatorname{var}(G_{i_1 i_2 \ldots i_k} \mid J = 0) = \frac{2}{k^3} C_{k,3} + \frac{1}{k^2} \frac{2}{k} C_{k,2} = \frac{k^2-1}{3k^2}; \quad (5.7)$$

risultato questo che è il massimo di (5.6), per un fissato n e sotto le condizioni (note) $\sum_1^k q_i = 1, q_i > 0$.

6. Impiego della distribuzione chi-quadrato. Caso generale.

Sotto l'ipotesi $J = 0$, la v.c.

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - nq_i)^2}{nq_i} \quad (6.1)$$

si distribuisce, al divergere di n, secondo una v.c. χ^2 centrale con $(k-1)$ gradi di libertà. Ricordando poi gli eventi precisati al paragrafo 4 e la (4.13), (4.14), (4.15), potremo affermare che l'evento

$$\{\bigcap A_{i_1 i_2 \ldots i_k}\}$$

individua in R^{k-1} un solido, che indicheremo con S_g, e in esso il valore della (4.1) sarà non superiore a g. Ci proponiamo ora di ricavare il più grande valore di X^2 contenuto in S_g. A tale scopo consideriamo la regione $(n_1, n_2, \ldots n_k) \in R_{i_1 i_2 \ldots i_k}$, in tal caso la faccia del solido S_g, in R^{k-1}, avrà l'equazione:

$$ng = \sum_{i=1}^{k-1} (\alpha_i - \alpha_k) n_i + n\alpha_k \quad (6.2)$$

dedotta dalla (4.10), ponendo in essa $G_{i_1 i_2 \ldots i_k} = g$.
Ricordando poi che:

$$n_k = n - \sum_{i=1}^{k-1} n_i \quad (6.3)$$

la (6.1) diventa:

$$X^2 = \sum_{i=1}^{k-1} \frac{(n_i - nq_i)^2}{nq_i} + \frac{\left(n - \sum_{i=1}^{k-1} n_i - nq_k\right)^2}{nq_k} . \qquad (6.4)$$

Per risolvere il problema prima precisato useremo il metodo dei moltiplicatori di Lagrange, e la funzione da estremare sarà quindi:

$$F = X^2 + \lambda \left[\sum_{i=1}^{k-1} (\alpha_i - \alpha_k) n_i + n \alpha_k - ng \right] . \qquad (6.5)$$

Con le consuete operazioni:

$$\frac{\partial F}{\partial n_i} = \frac{2(n_i - nq_i)}{nq_i} - \frac{2\left(n - \sum_{i=1}^{k-1} n_i - nq_k\right)}{nq_k} + \lambda(\alpha_i - \alpha_k), \qquad (i = 1, 2, \ldots, k-1)$$

$$\frac{\partial F}{\partial \lambda} = \sum_{i=1}^{k-1} (\alpha_i - \alpha_k) n_i + n \alpha_k - ng ,$$

abbiamo il seguente sistema:

$$2q_k n_i + 2q_i \sum_{i=i}^{k-1} n_i - 2nq_i + \lambda nq_i q_k (\alpha_i - \alpha_k) = 0, (i = 1,2,\ldots k-1), \qquad (6.6)$$

$$\sum_{i=1}^{k-1} (\alpha_i - \alpha_k) n_i + n \alpha_k - ng = 0. \qquad (6.7)$$

Applicando a (6.6) la sommatoria da 1 a (k−1) e ricordando che:

$$\sum_{i=1}^{k-1} q_i = 1 - q_k, \quad \sum_{i=1}^{k} \alpha_i q_i = 0 \text{ e pertanto } \sum_{i=1}^{k-1} \alpha_i q_i = - \alpha_k q_k,$$

si ottiene:

$$\sum_{i=1}^{k-1} n_i = n - nq_k + \frac{\lambda n\, q_k\, \alpha_k}{2}\,. \tag{6.8}$$

Sostituendo (6.8) in (6.6) si ha:

$$n_i = nq_i - \frac{\lambda n\, q_i\, \alpha_i}{2}\,, \quad (i = 1,2,\ldots k)\ . \tag{6.9}$$

Moltiplicando (6.9) per α_i e sommando poi da 1 a (k−1), cioè:

$$\sum_{i=1}^{k-1} \alpha_i n_i = n \sum_{i=1}^{k-1} q_i \alpha_i - \frac{\lambda n}{2} \sum_{i=1}^{k-1} q_i \alpha_i^2\,,$$

ne deriva:

$$\sum_{i=1}^{k-1} \alpha_i n_i = -\, nq_k\, \alpha_k - \frac{\lambda n d}{2} + \frac{\lambda n q_k \alpha_k^2}{2}\,, \tag{6.10}$$

ricordando la (4.23) e che, per (5.2), $\sum_{i=1}^{k} q_i \alpha_i^2 = d.$

Sostituendo ora in (6.7) la (6.10) e la (6.8) si ottiene:

$$\lambda = -\,\frac{2g}{d} \tag{6.11}$$

e pertanto le (6.9) diventano:

$$n_i = n\, q_i + \frac{n\, q_i\, \alpha_i}{d}\, g, \quad (i = 1, 2, \ldots, k). \tag{6.12}$$

Sostituendo quest'ultimo risultato in (6.1) avremo:

$$X^2 = \sum_{i=1}^{k} \frac{1}{nq_i} \frac{n^2 q_i^2 \alpha_i^2}{d^2} g^2 = \frac{n}{d^2} g^2 \sum_{i=1}^{k} \alpha_i^2 q_i = \frac{ng^2}{d}\,. \tag{6.13}$$

Si può concludere quindi che esiste un "ellissoide" tangente ai lati del "poliedro" S_g, caratterizzato dalla (6.4) in cui al posto di X^2 si è sostituito $\frac{ng^2}{d}$.

Indicando con $\chi^2_{\alpha,\ k-1}$ il valore critico della v.c. chi-quadrato con (k−1) gradi di libertà, è immediato notare che:

$$g^+_{\alpha,\ k-1} = \sqrt{\frac{d}{n}}\ \sqrt{\chi^2_{\alpha, k-1}} \tag{6.14}$$

è il maggiorante cercato di g_α, infatti $g^+_{\alpha,\ k-1}$ è tale che:

$$P\,[\,G > g^+_{\alpha,\ k-1} \mid J = 0\,] \leqslant \alpha\ .$$

Si può notare che, anche nel caso generale come nel caso $k = 3$, il maggiorante ora trovato differisce da quello trovato con la disuguaglianza di Bonferroni perchè in quest'ultimo al posto di $\Phi^{-1}\ (1 - \frac{\alpha}{k!})$ appare $\sqrt{\chi^2_{\alpha,\ k-1}}$.
Con l'aiuto delle tavole di E.S. Pearson e Hartlej [3] si è costruita la Tabella 1. che fornisce i valori di $\Phi^{-1}\ (1 - \frac{\alpha}{k!})$ e $\sqrt{X^2_{\alpha,\ k-1}}$, per alcuni valori di α e k. Da essa si ricava che per $k > 4$, i maggioranti ottenuti con la distribuzione del chi-quadrato sono minori di quelli ottenuti con la disuguaglianza del Bonferroni (unica eccezione il caso $\alpha = 0{,}01$), mentre per $k = 3$ sono quelli dedotti dalla disuguaglianza del Bonferroni ad essere minore degli altri (unica eccezione il caso $\alpha = 0{,}20$).

Tabella n.1

α	k = 3		k = 4		k = 5		k = 6		k = 7	
	$\Phi^{-1}(1-\frac{\alpha}{3!})$	$\sqrt{\chi^2_{\alpha,2}}$	$\Phi^{-1}(1-\frac{\alpha}{4!})$	$\sqrt{\chi^2_{\alpha,3}}$	$\Phi^{-1}(1-\frac{\alpha}{5!})$	$\sqrt{\chi^2_{\alpha,4}}$	$\Phi^{-1}(1-\frac{\alpha}{6!})$	$\sqrt{\chi^2_{\alpha,5}}$	$\Phi^{-1}(1-\frac{\alpha}{7!})$	$\sqrt{\chi^2_{\alpha,6}}$
0,20	1,8339	1,7943	2,3940	2,1547	2,9352	2,4472	3,4525	2,7001	3,9464	2,9255
0,10	2,1280	2,1459	2,6382	2,5002	3,1440	2,7891	3,6353	3,0391	4,1100	3,2626
0,05	2,3940	2,4477	2,8653	2,7954	3,3415	3,0802	3,8100	3,3272	4,2700	3,5484
0,02	2,7130	2,7973	3,1440	3,1366	3,5880	3,4174	4,0300	3,6601	4,4700	3,8776
0,01	2,9352	3,0348	3,3415	3,3682	3,7651	3,6437	4,1900	3,8841	4,6200	4,1002

7. Confronto fra i maggioranti ed i quantili asintotici esatti per k = 3 e k = 4

Per il caso $k = 3$, $k = 4$ e per $q_i = \frac{1}{k}$ $(i = 1,2, \dots, k)$, sempre sotto l'ipotesi $\Delta = 0$, è possibile confrontare i maggioranti $\sqrt{n}\ {}^{+}g_\alpha$ e $\sqrt{n}\ g_\alpha^{+}$ con i valori esatti $\sqrt{n}\ g_\alpha$ della distribuzione asintotica di G ricavati da Colombi [1] e da Brunazzo [2]. I confronti si faranno per i seguenti valori di α: 0,10; 0,05; 0,01.

Dalla Tabella 2 si nota che la differenza fra i valori maggioranti e quelli esatti è sempre inferiore al 5% dei valori di questi ultimi. Pertanto l'approssimazione data dai maggioranti può ritenersi soddisfacente.

Anche dal confronto, che verrà presentato in un prossimo lavoro, dei maggioranti con alcune distribuzioni campionarie reali ($n \leqslant 50$) di indici trasformabili nell'indice di concentrazione (1.1) si deduce che i maggioranti approssimano bene i quantili esatti anche per $k > 4$.

La presenza, nella Tabella 2, della relazione

$$\sqrt{n}\ g_\alpha = \frac{1}{k^2} \sqrt{k-1}\ \ h_\alpha$$

è giustificata dal fatto che, per comodità di calcolo, lo studio della distribuzione asintotica di G è stato sostituito, nel lavoro sopracitato dallo studio della distribuzione asintotica della variabile H, legata alla G dalla relazione:

$$\sqrt{n}\ G = \frac{1}{k^2} \sqrt{k-1}\ H.$$

Tabella n. 2

α	k	$\sqrt{d} = \sqrt{\frac{k^2-1}{3k^2}}$	$\sqrt{n}\ {}^{+}g_{\alpha,k} = \sqrt{d}\ \Phi^{-1}(1-\frac{\alpha}{k!})$	$\sqrt{n}\ g^{+}_{\alpha,k} = \sqrt{d}\ \sqrt{\chi^2_\alpha}$	$h_{\alpha.}$	$\sqrt{n}\ g_{\alpha.} = \frac{1}{k^2}\ \sqrt{k-1}\ h_\alpha$
0,10	3	0,5443	= 1,1582	= 1,1680	7,1149	= 1,1180
	4	0,5590	= 1,4747	= 1,3976	12,2785	= 1,3292
0,05	3	0,5443	= 1,3031	= 1,3323	8,1243	= 1,2766
	4	0,5590	= 1,6017	= 1,5626	13,75	= 1,4884
0,01	3	0,5443	= 1,5977	= 1,6519	10,10	= 1,5870
	4	0,5590	= 1,8679	= 1,8828	16,59	= 1,7959

BIBLIOGRAFIA

[1] R. COLOMBI (1980). Su un impiego della variabile casuale arcotangente incompleta nella verifica di una ipotesi relativa ai parametri di una trinomiale. *Atti della XXX Riunione Scientifica S.I.S.*, Vol. 2°

[2] A. BRUNAZZO Sulla distribuzione campionaria asintotica di un indice di concentrazione di Gini (Stesso quaderno pp.).

[3] E.S. PEARSON H.O. HARTLEY (1966). *Biometrika Tables for Statisticians*. Third Edition. Cambridge University Press.

RESUME

Les valeurs majorant de la distribution asynptotique d'un indice de concentration entre les classes de Gini

Dans ce travail on a calculé, avec deux differentes méthodes, les valeurs majorant des quantiles de la distribuion d'échantillons d'un indice de concentration de Gini. On a calculé les majorants dans l'hypothése d'un énchantillon de Bernoulli tiré d'une population sans concentration et dans l'hypothése qu'on connait les quotas du caractére transferable que dans la population sont à chaque classe. La premiére méthode utilise la classique inegalité de Bonferroni, l'autre utilise, par contre, la distribution de la variable aléatoire χ^2.

SUMMARY

The upper bounds of the asymptotic distribution of a Gini's concentration index among the classes

In this paper we use two different procedures to obtain upper bounds for the quantiles of the sampling distribution of a Gini's concentration index among the classes.

The upper bounds are deduced supposing a Bernoullian sample taken from a null-concentration population and supposing that the quotas of transferable character referring to single classes are known.

The first procedure employs the classical Bonferroni inequality, the second one employs the Chi-squared distribution.